

> Develop Predictive Models Using Your Business Expertise

Clementine is an integrated data mining workbench, popular worldwide with data miners and business analysts alike. Clementine supports the entire data mining process, enabling users to quickly access and prepare data for modeling, rapidly build and compare models, and then deploy models into real-time operational environments. Clementine seamlessly integrates data mining results with other business systems and processes, giving decision makers a clearer view of current conditions and insight into future events.

Clementine's highly visual user interface makes "train of thought" analysis possible. Clementine produces "streams," a visual map of your data mining steps, to automatically document the analytical process. By interacting with a stream, data miners can add business knowledge—the key to successful data mining—to the data mining process. The Clementine interface makes it easy to explore ideas or intuitive hunches. And because it allows you to focus on knowledge discovery, rather than on performing technical tasks such as writing code, you achieve results quickly.

Select from an unparalleled breadth of techniques

Clementine offers a full range of machine-learning and statistical techniques for clustering, classification, association, and prediction. SPSS Inc. has more than 35 years' experience in predictive analytics, and its algorithms are calibrated and verified to deliver powerful data mining results. With Clementine, you can be assured of having the analytical techniques you need to attain the most reliable results.

See more results, more quickly, with Clementine 8.5

Clementine 8.5 offers features that further increase the versatility this solution has always offered, improving your data mining productivity. Building on its existing, open, standards-based architecture, Clementine 8.5 enables organizations to further leverage their existing technology investments.

- Organizations using Microsoft® SQL Server™ 2000 Analysis Services can build Microsoft decision trees directly from the Clementine interface
- Companies using Sun™ Solaris™ UltraSPARC® servers will enjoy improved performance in processing large datasets, because Clementine 8.5 supports the 64-bit port on these servers
- Organizations can browse association models and automatically deploy other models built in Clementine into IBM® DB2® Data Warehouse Edition
- Customers of DataDistilleries B.V., which is now part of SPSS Inc., can author and automatically deploy streams from Clementine 8.5 into the predictive applications they currently use

To deepen the level of insight that can be gained from data mining, Clementine 8.5 adds a new association viewer. It also incorporates CARMA, a high-speed association detection algorithm that delivers not only faster processing but also more detailed information.

Integrate with existing information systems

With Clementine, there's good news for your IT staff. Clementine is an open, standards-based system that easily integrates with your organization's existing operational systems and databases. Clementine leverages your database investment by performing much of the data preprocessing tasks—typically 80-90 percent of your data mining effort—within the database. Because preprocessing occurs on the database tier, rather than on the server or client tiers, you save time, maximize resources, and minimize network traffic.

Put results to work in your organization

SPSS offers several deployment options to meet your needs for in-database or real-time scoring. Clementine exports not only the model but also all data mining steps—including data access, modeling, and post-processing—saving your organization time and money. Clementine also uses the industry-standard predictive model markup language (PMML), which means that language generated in Clementine can be easily embedded in other systems. As a result, you can efficiently publish processes or streams and transparently deliver results to other software applications.

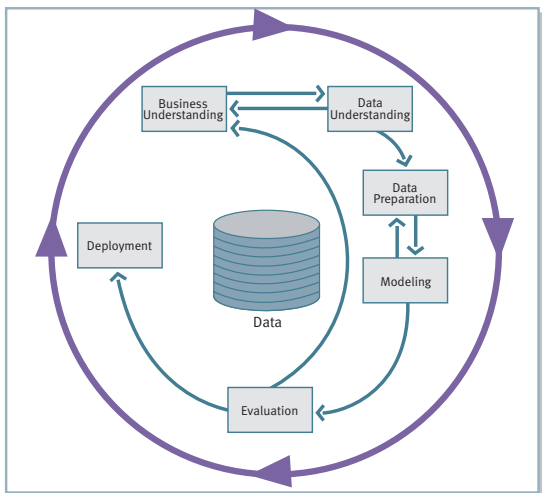


Features*

CRISP-DM

Help your company's analysts and data miners focus on business problem solving, rather than on the data mining technology. Clementine supports the CRoss-Industry Standard Process for Data Mining (CRISP-DM), the de facto industry standard, at every step of the process. Clementine includes an easy-to-use project manager that enables you to smoothly execute each step in your data mining projects using this methodology.

The diagram below shows the CRISP-DM process. Following the diagram is a summary of the capabilities of the Clementine solution that apply to each step of the CRISP-DM process, as well as a brief discussion of other features that make Clementine the data mining workbench of choice.



The CRISP-DM process, as shown above, enables data miners to efficiently implement data mining projects that yield measurable business results.

Business understanding

Clementine's visual interface makes it easy for your data miners to apply business knowledge to data mining projects. In addition, optional, business-specific Clementine Application Templates (CATs) are available that, when used in conjunction with Clementine, deliver data mining results faster. CATs support flat files as well as commonly used database layouts conforming to data warehouse standards. The CATs include the:

- CRM CAT**
- Telco CAT**
- Fraud CAT**
- Microarray CAT**

Data understanding

To provide a comprehensive "first look" at your data, Clementine enables you to automatically generate subsets of data from graphs and tables. Through the data audit node, you can create summary statistics, histograms, and distribution graphics for each data field and display them in an easy-to-read matrix.

- Visually interact with your data
 - Select a node or field and view information in a table
 - Create histograms, distributions, line plots, and point plots
 - Use Web association detection
 - Display 3-D, panel, and animated graphs

Data preparation

- Bring data into Clementine using a variety of methods and sources
 - Access structured (tabular) data
 - Use the security controls of your selected database or operating system to ensure data security
 - Access ODBC-compliant data sources. The included SPSS Data Access Pack enables you to work with data in Oracle®, Microsoft SQL Server, IBM DB2, Sybase®, and Informix® databases
 - Import delimited and fixed-width text; SPSS®, and SAS® 6,7,8, and 9 files

- Analyze unstructured (textual) data
 - Automatically extract concepts from documents and from text notes in databases, using Text Mining for Clementine**
- Analyze Web site data
 - Automatically extract Web site events from Web logs, using Web Mining for Clementine**
- Data output
 - Work with delimited and fixed-width text; ODBC; Microsoft Excel; SPSS; and SAS 6,7,8, and 9 files
- Choose from various data-cleaning options
 - Remove or replace invalid data
 - Automatically fill in missing values
- Manipulate data
 - Work with complete record and field operations, including:
 - Field filtering, naming, derivation, and value replacement
 - Record selection, sampling, merging, and concatenation; sorting, aggregation and balancing; deriving new fields based on conditional criteria; and calculating new fields
 - Specialized manipulations for showing the "history" of values and converting set variables into flag variables
 - Automatically bin data based on n-tiles, mean and standard deviation, rank, or a fixed width
 - Re-group or collapse categories and transform one set of discrete values to another
 - Customize the natural order of fields
- Merge records using any of the following types of joins:
 - Inner join
 - Full outer join
 - Partial outer join, both left and right
 - Anti-join (the opposite of an inner join)

Modeling

- Prediction and classification
 - Neural networks (multi-layer perceptrons trained using error-back propagation with momentum, radial basis function, and Kohonen networks)
 - Browse the importance of the predictors
 - Decision trees and rule induction [C5.0 and Classification and Regression Trees (C&RT)]
 - Browse and interact with decision trees
 - Browse, collapse, and expand decision rules
 - Linear regression, logistic regression, and multinomial logistic regression
 - View model equation and advanced statistical output
- Clustering and segmentation
 - Kohonen network, K-means, and TwoStep
 - View summary statistics and distributions for fields between clusters
- Association detection
 - GRI, Apriori, sequence
 - CARMA
 - Filter, sort, and subset association models with the new association model viewer
 - Use Apriori-, sequence-, and CARMA-generated models for scoring
- Data reduction
 - Factor analysis and principal components analysis
 - View model equation and advanced statistical output
- Combine models for greater accuracy through meta-modeling
 - Multiple models can be combined, or one model can be used directly for building a second model
- Import PMML-generated models created in other tools, such as AnswerTree® and SPSS for Windows®

- Clementine External Module Interface (CEMI) for custom algorithms
 - Add data access, modeling, and graphics
 - Create node palettes and generated model palettes, including PMML-generated models created in other tools, using custom CEMI icons
- An algorithm user manual is included, which explains the theories and methods behind the algorithms included in Clementine

Evaluation

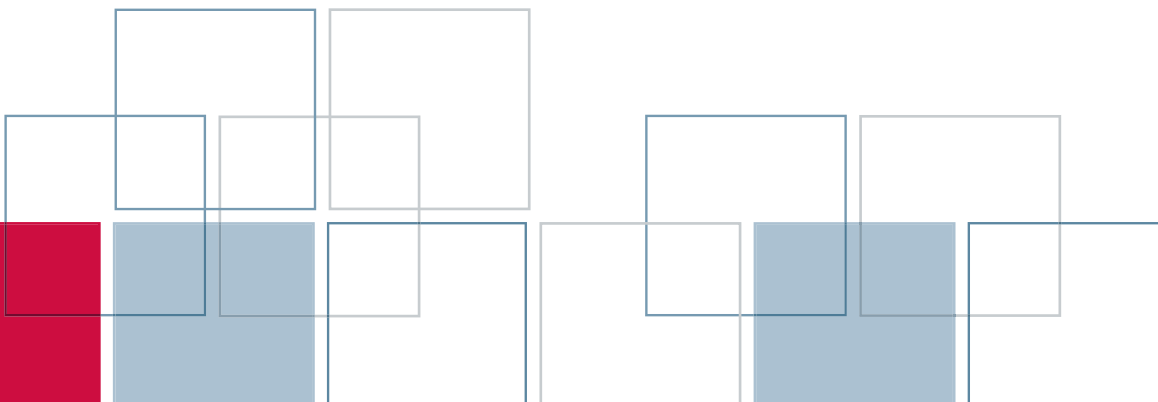
- Easily evaluate models using lift, gains, profit, and response graphs
 - Use a one-step process that shortens project time when evaluating multiple models
 - Define hit conditions and scoring expressions to interpret model performance

Deployment

Clementine offers a broad array of deployment capabilities to meet your organization's needs. Models built in Clementine can be directly deployed into other SPSS predictive applications and into other vendors' technologies.

- Clementine Solution Publisher (optional)
 - Automate the export of all operations, including data access; data manipulations; model scoring, including combinations of models; and post-processing
 - Work in a runtime environment for executing image files on target platforms
 - Easily update solutions through small image files
- Cleo™ (optional) a Web-based solution for rapid model deployment
 - Enable multiple users to simultaneously access and immediately score single records, multiple records, or an entire database, through a customizable browser-based interface

- Clementine Batch Mode
 - Automate production tasks, while working outside the user interface
 - Automate Clementine processes from other applications or scheduling systems
 - Generate encoded passwords
 - Launch commands using a command file
- Scripting
 - Automate command-line scripts or scripts associated with Clementine streams to automate repetitive tasks in the user interface. Scripts generally perform the same types of actions that otherwise are carried out using a mouse or keyboard.
 - Execute selected lines from a stream, SuperNode, or stand-alone script, using an icon on the toolbar
 - Update stream parameters within a standalone script
 - Refresh source nodes automatically upon stream execution
- Export generated models as PMML
- Perform in-database scoring, which eliminates the need for—and costs associated with—transferring data or performing calculations on the client
 - Generate SQL from linear regression and factor-generated models
 - Generate SQL from decision trees and rule sets
- Use the bulk loading capabilities of your database
 - Increase performance during data export by using a bulk loader program that's specific to your database. Fine-tune various options, including row-wise or column-wise binding for loading via ODBC, and batch-size setting for batch commits to the database.



Scalability and integration

Scalability

- Use in-database mining and server-tier processing to leverage parallel database implementations
- Minimize network traffic via intelligent field projection—Clementine pulls data only as needed from your data warehouse and passes only relevant results to the client

Integration

- Clementine integrates with SPSS and AnswerTree, and supports Cleo through a built-in Wizard
- It also integrates with other databases, market automation and software tools
- On Sun Solaris systems, Clementine 8.5 supports Kerberos security, a network authentication protocol, and runs as non-root
- Clementine 8.5 supports the export of streams to SPSS predictive applications that incorporate products formerly offered by DataDistilleries B.V.
- In addition, Clementine 8.5 integrates with:
 - Microsoft Analysis Services, for building decision trees
 - IBM DB2 Warehouse Edition, for deploying models to Intelligent Miner Scoring and the viewing of Clementine models with Intelligent Miner Visualization

System requirements

Clementine Client:

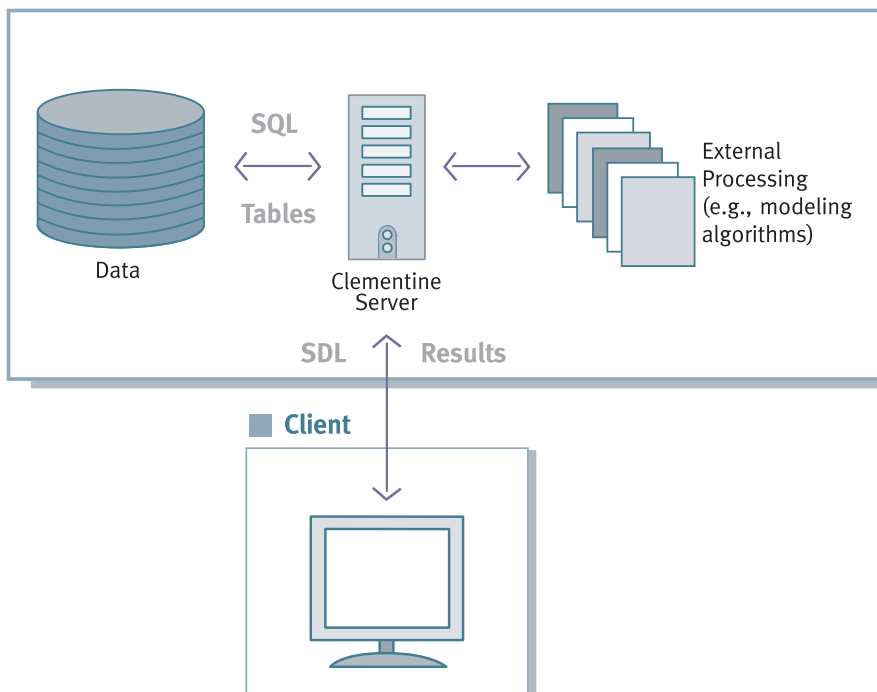
- Operating system: Microsoft Windows Me, Windows XP Home or Professional, Windows NT® 4.0 with Service Pack 6, Windows 2000 or 2003
- Memory: 256MB RAM recommended. 512MB recommended when using CATs or other large datasets.
- Minimum free disk space: 320MB
- A CD-ROM drive is required for installation
- Web browser: Internet Explorer version 5.x or later or Netscape® 6 or later, for running the online help system

Clementine Server, Clementine Solution Publisher Runtime (optional), and Clementine Batch Mode:

- Operating system: Windows 2000 Professional or Advanced Server, Windows 2003 Server, or Windows NT 4.0 with Service Pack 6 or later; Sun Solaris™ 2.7, 2.8, or 2.9; HP-UX 11i; IBM AIX® 4.3.3 or 5.2; or OS/400® (on the IBM eServer iSeries™) V5R2 with OS/400 Portable Applications Solution Environment (PASE, 5722-SS1 Option 33)

- Hardware: Pentium®-compatible processor, SPARC® for Solaris (Sun UltraSPARC® 2 hardware or better, if running on Solaris), HP Workstation for HP-UX (PA-RISC processor; HP 9000 server if running on HP-UX), IBM RS/6000® for AIX (PowerPC® 233MHz or better if running on AIX), or IBM iSeries server
- Memory: 256MB RAM minimum; 512MB recommended
- Minimum free drive space: 128MB required for installation. Additional free disk space is required to run the program (for temporary files). 1GB is recommended. For Clementine Solution Publisher Runtime, the minimum free disk space required to install the software is 64MB, plus at least twice the disk space of the amount of data to be processed.
- A network adapter running TCP/IP protocol
- A CD-ROM drive is required for installation

Server



Clementine uses a highly scalable, three-tiered architecture for data mining, as shown in this diagram. The Clementine Client tier (shown at the bottom of the illustration) passes stream description language (SDL) to Clementine Server. Clementine Server then analyzes these particular tasks to determine which tasks it can push to the database. After the database runs the tasks it can process, it passes only the relevant aggregated tables to Clementine Server. If you are using a CEMI, such as Text Mining for Clementine or Capri,™ Clementine Server passes the relevant tasks to that particular external process.

Cleo (optional):

- Web server: Cleo requires at least one server computer that meets the following minimum requirements. Using additional processors, faster processors, and more RAM will improve performance.
 - Operating system: Server versions of Windows 2000 or 2003, Windows NT 4.0 Server with Service Pack 5 (cannot be installed on Windows NT Terminal Server), or Sun Solaris Version 7 or later with UltraSPARC 2 hardware or better
 - Hardware: 500MHz processor or faster
 - Memory: 512MB RAM
 - Minimum free drive space: 700MB
 - A graphics adapter with 800 x 600 resolution (SVGA) or higher, capable of displaying at least 256 colors
 - A network adapter running the TCP/IP protocol
- Repository: the system requires a database to serve as a repository for published content, framework settings, and other information. The following databases are supported:
 - Microsoft SQL Server 2000
 - Oracle 8i, version 8.1.7
- Data warehouse: the system can be configured to access data from a data warehouse or database. The system has only been tested with SQL Server 2000 and Oracle 8i databases.
- Web client: content is delivered to clients as standard HTML pages. Supported browsers include:
 - Internet Explorer version 5.5 with Service Pack 2 or version 6.0 for Windows
 - Internet Explorer version 5.2 for Macintosh®
 - Netscape 6.2

Text Mining for Clementine (optional)*:**

Client version requirements

- Clementine 7.1 or later
- Operating system: Windows 2000, Windows 2000 Professional, or Windows NT 4.0 with Service Pack 6 or later
- Memory: 512MB RAM
- Minimum free disk space: 85MB, plus space for databases
- Web Browser: Internet Explorer 5.0 or later or Netscape 6.0 or later is required to use the Viewer node

Server version requirements

- Operating system: Server or Professional versions of Windows 2000 or 2003; Windows NT Server 4.0 with Service Pack 6 or later; or Solaris 8 or 9
- Memory: 256MB RAM
- Minimum free disk space: 85MB, plus space for databases

Web Mining for Clementine (optional):

Client version requirements

- Clementine 7.2 or later
- Operating system: Windows XP, or Server or Professional versions of Windows 2000 or 2003
- Memory: 512MB RAM
- Minimum free disk space: twice the amount of raw Web data being processed

Server version requirements

- Operating system: Windows Professional versions of Windows 2000 or 2003
- Memory: 512MB RAM
- Minimum free disk space: twice the amount of raw Web data being processed

Clementine Application Templates (optional):

- Clementine 7.2
- Memory: 512MB RAM recommended

Capri for Clementine (optional):

Plug-in algorithm for detecting sequences

- Operating system: Windows 98, 2000, 2003, Windows NT 4.0 with Service Pack 3 or later, or Sun Solaris 2.6
- Hardware: Pentium-compatible processor or higher for Windows, or SPARC for Solaris
- Memory: 128MB RAM
- Minimum free drive space: 3MB
- A CD-ROM drive is required for installation

*** Requirements for the Japanese-language version differ. See www.spss.com/lexiquet/systemrequirements.



To learn more, please visit www.spss.com. For SPSS office locations and telephone numbers, go to www.spss.com/worldwide.

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners.
© Copyright 2004 Integrated Solutions Ltd. CLM85SPC-0404